STA 431s17 Assignment Two[1]

1. Two latent explanatory variables X_1 and X_2 (say motivation and ability) potentially have non-zero covariance. Four job performance measures D_1, D_2, D_3 and D_4 are potentially related to X_1 and X_2 as follows:

$$
\begin{aligned}
D_1 &= \alpha_1 + \beta_{11}X_1 + \beta_{12}X_2 + \epsilon_1 \\
D_2 &= \alpha_2 + \beta_{21}X_1 + \beta_{22}X_2 + \epsilon_2 \\
D_3 &= \alpha_3 + \beta_{31}X_1 + \beta_{32}X_2 + \epsilon_3 \\
D_4 &= \alpha_4 + \beta_{41}X_1 + \beta_{42}X_2 + \epsilon_4,
\end{aligned}
$$

where the α and β quantities are unknown parameters, ϵ_1 through ϵ_4 are independent of one another and independent of X_1 and X_2, and everything is normally distributed.

   (a) Make a path diagram of this model. Write β_{ij} parameters on the appropriate arrows.

   (b) What are the unknown parameters of this model? I count 21. You will have to make up some notation for the expected values, variances and covariances.

2. Pigs are routinely given large doses of antibiotics even when they show no signs of illness, to protect their health under unsanitary conditions. Pigs were randomly assigned to one of three antibiotic drugs. Dressed weight (weight of the pig after slaughter and removal of head, intestines and skin) was the response variable. Explanatory variables are Drug type, Mother's live adult weight and Father's live adult weight. Data are in the plain text file

   `http://www.utstat.toronto.edu/~brunner/data/legal/pigweight.data.txt`.

   Once you have the raw data file open in a Web browser, you need to save the page to your computer and drag it to the `myfolders` sub-folder in your shared folder — that is, to the folder that is shared between your computer and the virtual `linux` machine on which SAS is installed. Exactly how you save a web page to your computer depends on your Web browser.

   - In Firefox, choose `Save Page As` from the `File` menu.
   - In Chrome, click on the wrench icon in the upper right corner, and choose `Save Page As`.
   - In Safari, choose `Save As ...` from the `File` menu, and then under `Format`, choose `Page Source`.

For this assignment, all you have to do is

(a) Use `proc freq` to make a frequency distribution of drug condition.

(b) Use `proc means` to calculate means and standard deviations for all the quantitative variables, separately for each drug condition.

(c) Use `proc corr` to produce a correlation matrix of the quantitative variables. By default you will also get means and standard deviations *not* broken down by country.

While it will be useful to use `kars.sas` (presented in lecture) as an example, please do not imitate the whole thing. Do only what you are asked to do. For example, you are *not* being asked to make dummy variables, and you are *not* being asked to do a regression.

Be able to answer questions like the following based on your results file:

- How many pigs received Drug 2?
- What percentage of pigs received Drug 2?
- What is the standard deviation of Father's Weight for pigs receiving Drug 3?
- What is the mean dressed weight of the pigs in the sample?
- What is the correlation between Mother's weight and Father's weight? What is the p-value? Is it statistically significant? Using the usual $\alpha = 0.05$ significance level, are you able to conclude that heavier fathers tend to be paired with lighter mothers?

Please follow these guidelines. Marks will be deducted if you do not.

- Put your name and student number in a `title` or `title2` statement.
- Do not write anything on the printouts except your name and student number. The other questions are just practice for the quiz, and are not to be handed in.
- Bring your log file to the quiz, *not* just a listing of the program file.
- The log file and the output file must be from the same run of SAS.
- You must use *your* installation of SAS, not the installation on someone else's computer.

3. For each of the following distributions, derive a general expression for the Maximum Likelihood Estimator (MLE); don't bother with the second derivative test. Then use the data to calculate a numerical estimate; you should bring a calculator to the quiz in case you have to do something like this. You will not be asked to carry out the second derivative test.

   (a) $p(x) = \theta(1-\theta)^x$ for $x = 0, 1, \ldots$, where $0 < \theta < 1$. Data: `4, 0, 1, 0, 1, 3, 2, 16, 3, 0, 4, 3, 6, 16, 0, 0, 1, 1, 6, 10.`

   (b) $f(x) = \frac{1}{\theta} e^{-x/\theta}$ for $x > 0$, where $\theta > 0$. Data: `0.28, 1.72, 0.08, 1.22, 1.86, 0.62, 2.44, 2.48, 2.96`

   (c) $f(x) = \frac{\alpha}{x^{\alpha+1}}$ for $x > 1$, where $\alpha > 0$. Data: `1.37, 2.89, 1.52, 1.77, 1.04, 2.71, 1.19, 1.13, 15.66, 1.43`

   (d) $f(x) = \theta x^{\theta - 1}$ for $0 < x < 1$, where $\theta > 0$. Data: `0.04, 0.69, 0.86, 0.24, 0.99`

4. For each of the distributions in Question 3, derive a formula for the Method of Moments estimator, and calculate it for the given data. To do this you need the expected values, and while it would be "interesting" to calculate them yourself, that's not a goal of this course. So, here are the expected values.

   (a) Geometric: $E(X) = \frac{1-\theta}{\theta}$

   (b) Exponential: $E(X) = \theta$

   (c) Pareto: $E(X) = \frac{\alpha}{\alpha - 1}$ for $\alpha > 1$. For $0 < \alpha \leq 1$, the expected value does not exist.

   (d) Beta with $\alpha = \theta$ and $\beta = 1$: $E(X) = \frac{\theta}{\theta + 1}$.

5. Let $Y_i = \beta x_i + \epsilon_i$ for $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are a random sample from a normal distribution with expected value zero and variance σ^2. The parameters β and σ^2 are unknown constants. The numbers $x_1, \ldots, x_n$ are known, observed constants.

   (a) What is the parameter space Θ?

   (b) Find the Maximum Likelihood Estimator of the pair (β, σ^2). Show your work.

   (c) Based on the small data set below, calculate the MLE for β. Your answer is a number. Bring a calculator in case you have to do something like this on the quiz.

```
x  0.0  1.3  3.2 -2.5 -4.6 -1.6  4.5  3.8
y -0.8 -1.3  7.4 -5.2 -6.5 -4.9  9.9  7.2
```